Intuitive diffraction model for multistaged optical systems

Eric L. Shirley

A simplified framework is motivated in which many diffraction effects can be treated, especially in multistaged optical systems. The results should be especially helpful for short wavelengths and broadband sources, for which numerical calculations can be most difficult.

OCIS codes: 050.1960, 120.5630.

1. Introduction and Motivation

Diffraction modifies the flow of light through optical systems, changing it from what is expected according to geometrical optics. To fully characterize the performance of an optical system, it can be necessary to account for diffraction effects, especially for purposes of accurate radiometry. However, diffraction effects can be exceedingly complicated and virtually impossible to describe exactly, especially in multistage optical systems. Therefore the merit of a description of diffraction effects, or perhaps a preliminary description, might be a combination of its quantitative accuracy and its conceptual simplicity. In this paper an intuitive framework is laid out that permits one to estimate diffraction effects on the power received by a detector in a blackbody calibration, but diffraction effects in many analogous optical systems can also be considered with suitable adaptation of the same framework.

Many calibrations of blackbodies, in which one traditionally calibrates the radiance temperature of a blackbody radiometrically, use an optical setup like the one shown in Fig. 1. A blackbody cavity emits radiation through a circular core opening of radius R_{BB}. This opening is at a distance d_{BB} from the source defining aperture with radius R_s, which defines the portion of the blackbody core viewed by the detector. Because the detector may sense other radiation in addition to radiation emitted by the blackbody, one or more nonlimiting apertures or baffles can be placed between the blackbody and the detector to reduce such stray light.

E. Shirley (eric.shirley@nist.gov) is with the Optical Technology Division, National Institute of Standards and Technology, 100 Bureau Drive, Mail Stop 8441, Gaithersburg, Maryland 20899-8441.

Received 16 July 2003; revised manuscript received 31 October 2003; accepted 6 November 2003.

Ideally, the radiance temperature and source defining aperture area are transferable characteristics of a blackbody that do not change when the blackbody is taken from one optical setup to another. In particular, one usually desires that these properties of the blackbody do not change between when it is calibrated and when it is used as a calibrated source. Diffraction effects specific to each optical setup do change, however. In a blackbody calibration of the type described here, diffraction losses can occur because the source defining aperture diffracts light and so prevents some of it from reaching the detector. Conversely, some light that should not reach the detector aperture may accidentally reach it because of diffraction by the nonlimiting apertures. To determine the transferable blackbody characteristics most accurately, the power that is measured during a calibration should be interpreted with all of the above diffraction effects taken into account.

These issues have been discussed for some time, as considered, for example, by Blevin,[1] Steel *et al.*,[2] Boivin,[3] and Shirley.[4,5] In all these studies, the authors treated diffraction effects in optical setups by considering diffraction by apertures throughout the setup one aperture at a time. In principle, it is better to describe end-to-end propagation of light through the entire optical setup in a coherent fashion.[6] However, this can be numerically intensive, especially at small wavelengths. Moreover, the small-wavelength behavior of diffraction effects on spectral power reaching the detector can follow simple trends (e.g., giving a relative contribution to spectral power that is proportional to wavelength λ), and one can often treat diffraction using simplified formulas. If the overall diffraction effects are sufficiently small, simplified formulas may be adequate.

In this paper I present an intuitive framework

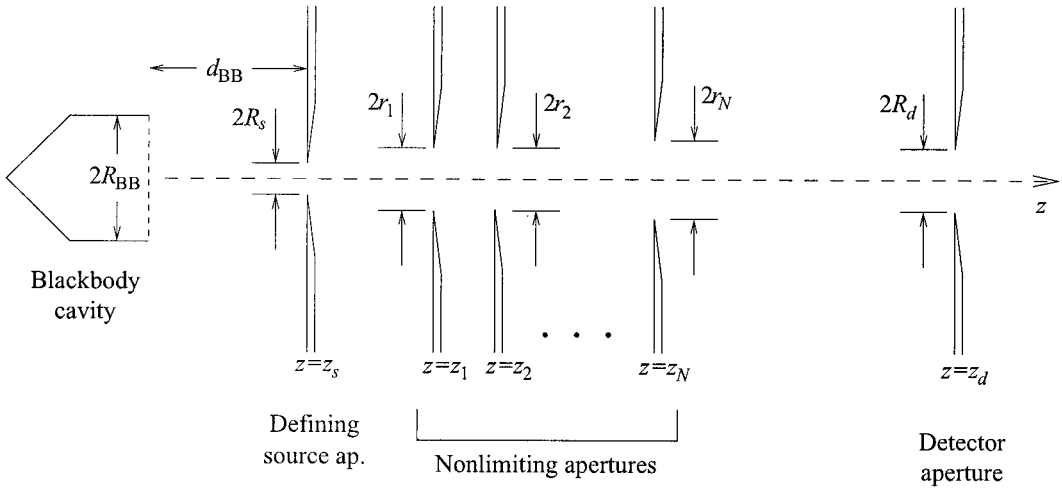

Fig. 1. Schematic optical setup for blackbody calibration. Cylindrical symmetry with respect to the optical (z) axis is assumed. Shown are the blackbody cavity opening, the source defining aperture (ap.), nonlimiting apertures to prevent stray light, and the detector aperture with relevant dimensions and distances.

for estimating all the above diffraction effects efficiently, hopefully with minor approximation, emphasizing the possibility of treating small wavelengths, extended sources (e.g., large source defining apertures), and broadband sources. In part, this follows from a lengthy consideration of and experience in the modeling of how diffraction affects the throughput of an optical system. In addition to diffraction losses due to the defining aperture and gains due to nonlimiting apertures, in this paper I also account for some subtler diffraction effects. Namely, the diffraction effects of nonlimiting apertures can be affected because other optical elements can geometrically obstruct light scattered by, or to be scattered from, their edges and because the source's illumination of their light-diffracting edges can be reduced by diffraction when the source defining aperture is sufficiently small.

The framework presented here relies on several assumptions that are often valid. The most important assumptions are that we are considering the propagation of polychromatic light with a sufficiently small characteristic wavelength and that apertures and other optical elements are either well overfilled or underfilled so that their perimeters are sufficiently far from geometric shadow boundaries or penumbra regions. The framework works well in certain limits and should not work in other situations. Results found within the framework are obtained for optical setups similar to those used in actual blackbody calibrations and are compared with numerically intensive calculations that simulate end-to-end light propagation through the setup. These results indicate the plausibility of one using the framework in similar situations. In the examples considered, the required computer time was reduced by factors of approximately 100 or greater with the present framework as compared with the full numerical calculations.

2. Geometries Considered

We consider light propagating through a cylindrically symmetrical optical system depicted in Fig. 1. The system consists of a source, a series of N nonlimiting apertures, and a detector whose circular entrance aperture has radius R_d. We define the optical axis to be the z axis, with light moving mainly in the positive z direction but with small deflections in the x and y directions. The source may be a point source, an extended-area circular source with radius R_s, or a source in which radiation is generated behind a source defining aperture with radius R_s, as depicted in Fig. 1. In the former two cases, the point or extended-area source would be in the $z = z_s$ plane, centered with respect to the optical axis. In general, light must either be emitted at or pass through point $\mathbf{r}_s = (x_s, y_s, z_s)$, which can be anywhere on the source area or source aperture area for the latter two types of source. If the light reaches any point $\mathbf{r}_d = (x_d, y_d, z_d)$ on the detector aperture, it contributes to the power reaching the detector. The N apertures are called nonlimiting because they do not block the line of sight between any pair of points $\mathbf{r}_s$ and $\mathbf{r}_d$. A nonlimiting aperture i, for $i = 1$ to N, can be specified by the z coordinate of the aperture's plane z_i and the aperture radius r_i.

3. Development of the Framework

The objective of this paper is an efficient, approximate scheme to determine the spectral power incident on the detector aperture, including all diffraction effects. We first consider the simplest problem, which is to determine the irradiance for a single point source. We next consider the irradiance for the case of a circularly symmetrical extended-area source. Last, we consider the changes resulting from having radiation actually emitted behind the source defining aperture.

A. Diffraction Effects for a Point Source

For a monochromatic point source, the scalar radiation field at point $\mathbf{r}_d$ is given by

$$U(\mathbf{r}_d, \mathbf{r}_s) = U_G(\mathbf{r}_d, \mathbf{r}_s) + U_D(\mathbf{r}_d, \mathbf{r}_s) = \frac{\exp(ik|\mathbf{r}_d - \mathbf{r}_s|)}{|\mathbf{r}_d - \mathbf{r}_s|} + U_D(\mathbf{r}_d, \mathbf{r}_s). \quad (1)$$

The first term has the form of an outgoing free spherical wave emitted at point $\mathbf{r}_s$, and the second term is a complicated expression with indicated dependences and an implicit dependence on angular wave number $k = 2\pi/\lambda$. The second term accounts for diffraction effects, and the irradiance at $\mathbf{r}_d$ is affected by diffraction in approximately the same way as $|U(\mathbf{r}_d, \mathbf{r}_s)|^2$. Within the Kirchhoff diffraction theory by use of the Fresnel approximation, one obtains

$$U(\mathbf{r}_d, \mathbf{r}_s) \approx (i\lambda)^{-N} \int_{\text{aper } 1} \mathrm{d}^2\mathbf{r}_1 \ldots \int_{\text{aper } N} \times \mathrm{d}^2\mathbf{r}_N \frac{\exp[ik(|\mathbf{r}_s - \mathbf{r}_1| + \cdots + |\mathbf{r}_N - \mathbf{r}_d|)]}{(z_s - z_1)\ldots(z_d - z_N)}. \quad (2)$$

Here a distance is approximated, in a denominator that can be taken outside the integral, as the difference in z coordinates, but when determining the complex phase associated with a path length, we approximate a distance according to the pattern

$$|\mathbf{r} - \mathbf{r}'| \approx |z - z'| + \frac{(x - x')^2 + (y - y')^2}{2|z - z'|}. \quad (3)$$

Evaluating the total $U(\mathbf{r}_d, \mathbf{r}_s)$ is possible but can be numerically intensive, and prohibitively so at small wavelength λ. Furthermore, diffraction effects are smallest at the smallest wavelengths, making it even less helpful to expend considerable time and computational resources to determine such effects. Instead, I now provide an intuitive motivation for an approximate evaluation of the relative diffraction effects on the spectral irradiance. We have

$$\frac{E_\lambda(\mathbf{r}_d, \mathbf{r}_s; \lambda)}{E_{0,\lambda}(\mathbf{r}_d, \mathbf{r}_s; \lambda)} - 1 = \frac{|U_D(\mathbf{r}_d, \mathbf{r}_s)|^2}{|U_G(\mathbf{r}_d, \mathbf{r}_s)|^2} + 2\,\mathrm{Re}\left[\frac{U_G^*(\mathbf{r}_d, \mathbf{r}_s) U_D(\mathbf{r}_d, \mathbf{r}_s)}{|U_G(\mathbf{r}_d, \mathbf{r}_s)|^2}\right]. \quad (4)$$

Here $E_\lambda(\mathbf{r}_d, \mathbf{r}_s; \lambda)$ denotes the actual spectral irradiance at point $\mathbf{r}_d$ because of a point source located at point $\mathbf{r}_s$, including diffraction effects, whereas $E_{0,\lambda}(\mathbf{r}_d, \mathbf{r}_s; \lambda)$ denotes the analogous spectral irradiance expected according to geometrical optics, i.e., neglecting diffraction effects. The first term on the right-hand side is always positive and contributes to the spectral power everywhere on the detector aperture at all wavelengths. The second term is a cross term involving $U_G(\mathbf{r}_d, \mathbf{r}_s)$ and $U_D(\mathbf{r}_d, \mathbf{r}_s)$ and is oscillatory in at least two ways. First, it oscillates as a function of $\mathbf{r}_d$, and when it is averaged over an extended area of the detector aperture, the second term is largely self-canceling. This is especially true at small λ, where spatial oscillations are most rapid. The second term also oscillates as a function of λ, so that its effect on total power for a broadband source (such as a blackbody) can be small, again especially around small λ. From this point forward, we ignore the second term. Next, when we are considering diffraction effects for radiation from a point source, it is useful to first consider diffraction effects of a single nonlimiting aperture and then to consider diffraction effects of several nonlimiting apertures.

B. Single Nonlimiting Aperture

In the case of a single nonlimiting aperture, the boundary diffraction wave formulation[7,8] of the Kirchhoff diffraction theory yields

$$U_D(\mathbf{r}_d, \mathbf{r}_s) = \int_\Gamma \frac{\mathrm{d}\mathbf{l} \cdot (\mathbf{s}_d \times \mathbf{s}_s)}{4\pi(s_d s_s + \mathbf{s}_d \cdot \mathbf{s}_s)} \left\{\frac{\exp[ik(s_s + s_d)]}{s_s s_d}\right\}. \quad (5)$$

The line integral with respect to $\mathbf{l}$ is done around the aperture perimeter Γ in the right-hand direction about the direction of forward propagation, and we introduced $\mathbf{s}_d = \mathbf{l} - \mathbf{r}_d$ and $\mathbf{s}_s = \mathbf{l} - \mathbf{r}_s$. This replaces Kirchhoff's double integral with a single integral. As written, Eq. (5) is valid if the line of sight between $\mathbf{r}_s$ and $\mathbf{r}_d$ passes through the aperture, which is always true for a nonlimiting aperture. This suggests the notion that there is a physical meaning in the integrand related to scattering of light from each point on the aperture perimeter. Such a connection is actually not so straightforward, although $U_D(\mathbf{r}_d, \mathbf{r}_s)$ must somehow originate from light interacting with the aperture's material near its perimeter.

It is convenient to approximate the above line integral in the following fashion. Let $d_s = z_1 - z_s$ and $d_d = z_d - z_1$ denote the distances along the z axis between the source and the aperture and between the aperture and the detector planes, respectively. In the Fresnel approximation, we have a total path length related to the complex phase in Eq. (5) given by

$$s_s + s_d \approx d_s + d_d + \frac{(x_s - x_d)^2 + (y_s - y_d)^2}{2(d_s + d_d)} + \left(\frac{1}{2d_s} + \frac{1}{2d_d}\right)[(x - x_m)^2 + (y - y_m)^2]. \quad (6)$$

$$(x_m, y_m, z_1) = \left(\frac{x_s d_d + x_d d_s}{d_s + d_d}, \frac{y_s d_d + y_d d_s}{d_s + d_d}, z_1\right) \quad (7)$$

is where the line segment between $\mathbf{r}_s$ and $\mathbf{r}_d$ passes through the aperture plane. Only the last term on the right-hand side in approximation (6) depends on

$\mathbf{l} = (x, y, z_1)$, and this term depends on $\mathbf{r}_s$ and $\mathbf{r}_d$ only through x_m and y_m. It follows that we have

$$|U_D(\mathbf{r}_d, \mathbf{r}_s)|^2 \approx |U_D[(r_m, 0, z_d), (r_m, 0, z_s)]|^2, \quad (8)$$

with $r_m = (x_m{}^2 + y_m{}^2)^{1/2}$. In the Fresnel approximation, the right-hand side expression is

$$\mathbf{s}_d \times \mathbf{s}_s \approx [y(d_s + d_d), -(x - r_m) \times (d_s + d_d), 0], \quad (9)$$

$$s_d s_s + \mathbf{s}_d \cdot \mathbf{s}_s \approx \frac{(d_s + d_d)^2[(x - r_m)^2 + y^2]}{2d_s d_d}. \quad (10)$$

For sufficiently large r_m, one can use the method of stationary phase in the small-λ limit to perform the integration in Eq. (5), because the integral is dominated by the two sections of the aperture perimeter where the phase of the integrand is stationary, near $\mathbf{l} = (\pm r_1, 0, z_1)$, where we have

$$s_s + s_d \approx d_s + d_d + \frac{(x_s - x_d)^2 + (y_s - y_d)^2}{2(d_s + d_d)} + \left(\frac{1}{2d_s} + \frac{1}{2d_d}\right)\left[(\pm r_1 - r_m)^2 \pm \left(\frac{r_m}{r_1}\right)y^2\right], \quad (11)$$

$$\mathrm{d}\mathbf{l} \cdot (\mathbf{s}_d \times \mathbf{s}_s) \approx -[r_1 - (\pm r_m)](d_s + d_d)\mathrm{d}y, \quad (12)$$

$$s_d s_s + \mathbf{s}_d \cdot \mathbf{s}_s \approx \frac{(d_s + d_d)^2[(\pm r_1 - r_m)^2 + y^2]}{2d_s d_d}. \quad (13)$$

We obtain

$$U_D[(r_m, 0, z_d), (r_m, 0, z_s)] \approx \frac{I_+ + I_-}{4\pi d_s d_d}, \quad (14)$$

where, suppressing their arguments, we have

$$I_\pm = \frac{2d_s d_d \exp(i\phi_\pm)}{(d_s + d_d)[r_1 - (\pm r_m)]}\left[\frac{2\pi r_1 d_s d_d}{k r_m (d_s + d_d)}\right]^{1/2}. \quad (15)$$

The phases $\phi_\pm$, which have not been specified, vary relatively, so that, if one only seeks $|U_D(\mathbf{r}_d, \mathbf{r}_s)|^2$ averaged over the detector area or λ, it suffices to consider only the squared complex moduli of I_+ and I_-, yielding

$$|U_D(\mathbf{r}_d, \mathbf{r}_s)|^2 \approx \frac{r_1 d_s d_d}{2\pi k r_m (d_s + d_d)^3}\left[\frac{1}{(r_1 - r_m)^2} + \frac{1}{(r_1 + r_m)^2}\right], \quad (16)$$

where the omitted terms are approximated as being self-canceling. Use of the stationary-phase approximation to estimate asymptotic properties of the boundary diffraction wave is an obvious idea often anticipated by earlier research, perhaps most clearly by Keller in connection with the geometric theory of diffraction.[9] With this theory we can model diffraction by considering rays that scatter off edges only at discrete points and in certain distributions of directions arrived at by physical reasoning. The scattering efficiency is known to within the square of a diffraction coefficient that must be deduced from other diffraction calculations. Not surprising, the points $\mathbf{l} = (\pm r_1, 0, z_1)$ in this paper correspond to the points identified by Keller.

However, r_m can also be small or even zero, leading to a divergence in the above result and a loss of special meaning for the points $\mathbf{l} = (\pm r_1, 0, z_1)$. This divergent behavior is associated with a breakdown of the stationary-phase method when the asymptotic conditions of validity are not met. However, in practice, the divergent behavior will often give rise to only a mild degree of approximation. This is because, when we integrate over $\mathbf{r}_d$ for a fixed $\mathbf{r}_s$, or vice versa, a given value of r_m is sampled with a weight that is linear in r_m in the limit of $r_m \to 0$. Therefore the integrated result has only a finite error, which decreases to zero in the small-λ limit. In practice, the significance of this error is indicated by the degree of accuracy of results obtained by the present method.

We now make the abbreviations $u = k r_1{}^2(d_s{}^{-1} + d_d{}^{-1})$, $v = k r_1 r_m(d_s{}^{-1} + d_d{}^{-1})$, and $w = v/u = r_m/r_1$. In what follows, the results are expressed most succinctly in terms of v and w, whereas u and v correspond most closely to the arguments of Lommel functions, which are indicated by the same symbols, in discussions such as the one by Born and Wolf.[10] We obtain

$$\frac{|U_D(\mathbf{r}_d, \mathbf{r}_s)|^2}{|U_G(\mathbf{r}_d, \mathbf{r}_s)|^2} \approx \frac{1}{2\pi v(1 + v/u)^2} + \frac{1}{2\pi v(1 - v/u)^2} = \frac{1}{2\pi v(1 + w)^2} + \frac{1}{2\pi v(1 - w)^2}, \quad (17)$$

again presuming that the omitted terms can be neglected. The first and second terms shown arise from portions of the aperture perimeter that are farthest from and nearest to the point (x_m, y_m, z_1), respectively. Ultimately, we are considering diffraction effects that depend on u and v, which in turn depend only on λ, r_1, r_m, d_s, and d_d, some of which in turn depend on $\mathbf{r}_d$ and $\mathbf{r}_s$. In what follows, it is convenient to use the shorthand $N(\mathbf{r}_d, \mathbf{r}_s) = 1/[2\pi v(1 - w)^2]$ and $F(\mathbf{r}_d, \mathbf{r}_s) = 1/[2\pi v(1 + w)^2]$.

C. Multiple Nonlimiting Apertures

In the case of multiple nonlimiting apertures, two further conditions can still yield a convenient expression for important contributions to $|U_D(\mathbf{r}_d, \mathbf{r}_s)|^2/|U_G(\mathbf{r}_d, \mathbf{r}_s)|^2$. First, note that the factor of $\exp[ik(s_s + s_d)]/(s_s s_d)$ in the integrand in Eq. (5) is the product of two functions, $\exp(iks_s)/s_s$ and $\exp(iks_d)/s_d$. Correspondingly, in the case of a single aperture, light propagates as an unperturbed wave between the source plane and the aperture plane and between the aperture plane and the detector plane. This is no longer the case when there are multiple apertures, because light passing from the source plane to an

aperture plane or from that aperture plane to the detector plane will have to pass through other apertures. Treating diffraction by a given aperture as above accounts for all contributions of that aperture to first-order diffraction effects, provided that the geometric blocking effects of other apertures, when relevant, are considered. Second, for sufficiently broadband light, one can assume that diffraction effects of the apertures combine additively, so that, on the average, interference effects involving different apertures' contributions to $|U(\mathbf{r}_d, \mathbf{r}_s)|^2$ are largely self-canceling.

If these conditions are met, we hope to generalize the above functions $N(\mathbf{r}_d, \mathbf{r}_s)$ and $F(\mathbf{r}_d, \mathbf{r}_s)$ to a set of functions appropriate for each aperture i, $[N_i(\mathbf{r}_d, \mathbf{r}_s)]$ and $[F_i(\mathbf{r}_d, \mathbf{r}_s)]$. A function $N_i(\mathbf{r}_d, \mathbf{r}_s)$ would be calculated in the same way as was the function $N(\mathbf{r}_d, \mathbf{r}_s)$ if the point on the aperture closest to the appropriate point $(x_{m,i}, y_{m,i}, z_i)$ has unblocked lines of sight to both $\mathbf{r}_s$ and $\mathbf{r}_d$. However, the function $N_i(\mathbf{r}_d, \mathbf{r}_s)$ is zero otherwise. Likewise, a function $F_i(\mathbf{r}_d, \mathbf{r}_s)$ would be calculated in the same way as was the function $F(\mathbf{r}_d, \mathbf{r}_s)$ if the point on the aperture farthest from the appropriate point $(x_{m,i}, y_{m,i}, z_i)$ has unblocked lines of sight to both $\mathbf{r}_s$ and $\mathbf{r}_d$. However, the function $F_i(\mathbf{r}_d, \mathbf{r}_s)$ is zero otherwise. One then obtains the result, in the multiple-aperture case,

$$\frac{|U_D(\mathbf{r}_d, \mathbf{r}_s)|^2}{|U_G(\mathbf{r}_d, \mathbf{r}_s)|^2} \approx \sum_{i=1}^{N} [N_i(\mathbf{r}_d, \mathbf{r}_s) + F_i(\mathbf{r}_d, \mathbf{r}_s)] + \cdots, \tag{18}$$

where terms not listed are assumed to be small or nearly self-canceling.

D. Diffraction Effects for an Extended-Area Source

An extended-area source of the type we consider is equivalent to a set of mutually incoherent point sources distributed equally everywhere on the source aperture area. We can find the diffraction effects on the irradiance at point $\mathbf{r}_d$ by averaging the diffraction effects for all such point sources. Therefore, for the case of an extended-area source, we obtain

$$\frac{E_\lambda(\mathbf{r}_d)}{E_{0,\lambda}(\mathbf{r}_d)} - 1 = \frac{1}{\pi {R_s}^2} \int_{\text{source aper}} \mathrm{d}^2\mathbf{r}_s \frac{|U_D(\mathbf{r}_d, \mathbf{r}_s)|^2}{|U_G(\mathbf{r}_d, \mathbf{r}_s)|^2}. \tag{19}$$

As noted above, the approximate functions $F(\mathbf{r}_d, \mathbf{r}_s)$ and $N(\mathbf{r}_d, \mathbf{r}_s)$ diverge as $1/r_m$ nears $r_m = 0$, but the integration in Eq. (19) samples different values of r_m in a fashion that renders a convergent result. It is therefore wise, when we perform the integration in Eq. (19) numerically, to do it in a way that explicitly realizes this cancellation of the divergence. In this paper integration is performed in cylindrical polar coordinates in the source aperture plane with the origin placed at the point corresponding to $r_m = 0$.

E. Diffraction Effects for Radiation Originating Behind the Source Defining Aperture

The developments above were made with a point source or extended-area source in mind. These models for the source are idealizations for many real sources, such as blackbodies, where radiation is actually incident on a source defining aperture because of a radiating cavity that is behind the aperture. In geometrical optics, this would be a moot point, because the defining aperture would appear the same to the detector as an extended-area source, if the radiating entities in the cavity fill the field of view seen through the aperture.

However, there are three possible effects of the source defining aperture that concern us here. First, diffraction losses at the source defining aperture can prevent some flux from reaching the detector, implying that corrections are needed when source radiance, source aperture area, and measured power are related. These losses are well characterized and discussed elsewhere.[1–5] Second and third, the source defining aperture can affect diffraction effects of the N nonlimiting apertures by affecting the light that reaches their perimeters by geometrically blocking the light and by diffraction effects. To account for this, the functions $F_i(\mathbf{r}_d, \mathbf{r}_s)$ and $N_i(\mathbf{r}_d, \mathbf{r}_s)$, or the integrals thereof that quantify diffraction effects on power reaching the detector, can be scaled or evaluated slightly differently.

The diffraction effects of the source defining aperture and nonlimiting apertures combine to affect the power reaching the detector. We can characterize this effect at each wavelength by the ratio of the spectral power reaching the detector, $\Phi_\lambda(\lambda)$, to the spectral power that would reach the detector in geometrical optics, $\Phi_{0,\lambda}(\lambda)$. Calling this ratio $F(\lambda)$, and the difference between $F(\lambda)$ and unity $\epsilon(\lambda)$, we obtain

$$\Phi_\lambda(\lambda)/\Phi_{0,\lambda}(\lambda) = F(\lambda) = 1 + \epsilon(\lambda). \tag{20}$$

In the case of there being no nonlimiting apertures, we obtain

$$[\Phi_\lambda(\lambda)/\Phi_{0,\lambda}(\lambda)]|_{N=0} = F_L(\lambda) = 1 + \epsilon_L(\lambda), \tag{21}$$

a function that is characterized elsewhere. In the case of an extended-area (EA) source, we obtain

$$[\Phi_\lambda(\lambda)/\Phi_{0,\lambda}(\lambda)]|_{\text{EA}} = F_G(\lambda) = 1 + \epsilon_G(\lambda), \tag{22}$$

with

$$\epsilon_G(\lambda) = (\pi {R_\text{d}}^2)^{-1} \int_{\text{det. aper.}} \mathrm{d}^2\mathbf{r}_d [E_\lambda(\mathbf{r}_d)/E_{0,\lambda}(\mathbf{r}_d) - 1]. \tag{23}$$

Because this has identifiable contributions from each nonlimiting aperture, we can decompose this into each nonlimiting aperture's contribution:

$$\epsilon_G(\lambda) = \sum_{i=1}^{N} \epsilon_{G,i}(\lambda). \tag{24}$$

Note that we usually have $\epsilon_L(\lambda) < 0$ and $\epsilon_G(\lambda) > 0$.

In the case of a source with light originating behind the source defining aperture, we obtain

$$\Phi_\lambda(\lambda)/\Phi_{0,\lambda}(\lambda) = F(\lambda) = 1 + \epsilon_L(\lambda) + \alpha_G(\lambda), \tag{25}$$

Table 1. Specifications of the Optical Setup Used by Boivin

Aperture	Radius (mm)	z Coordinate (mm)
Extended-area source	$R_s = 0.5$	$z_s = 0$
1	$r_1 = 3.5$	$z_1 = 500$
2	$r_2 = 3.5$	$z_2 = 650$
3	$r_3 = 3.5$	$z_3 = 800$
4	$r_4 = 3.5$	$z_4 = 950$
Detector	$R_d = 1.25$	$z_d = 1350$

with

$$\alpha_G(\lambda) = \sum_{i=1}^{N} \alpha_{G,i}(\lambda). \tag{26}$$

Here, a function $\alpha_{G,i}(\lambda)$ replaces $\epsilon_{G,i}(\lambda)$. One either evaluates $\alpha_{G,i}(\lambda)$ as described below or one evaluates the function $\epsilon_{G,i}(\lambda)$ as described above and scales it according to a rule, $\alpha_{G,i}(\lambda) = w_i\epsilon_{G,i}(\lambda)$. In the latter case, w_i remains to be determined as described below. The algorithm for deciding which of the last two options to take can be stated as follows. As the first step, if we consider a point $\mathbf{r}_i$ on the perimeter of nonlimiting aperture i, the source defining aperture provides $\mathbf{r}_i$ a field of view into the region of space behind the $z = z_s$ plane. If and only if the radiating entities of the source fill this field of view, the perimeter of nonlimiting aperture i is fully illuminated.

If the perimeter of nonlimiting aperture i is not fully illuminated, the source defining aperture blocks some light from reaching the perimeter for purely geometric reasons. In this case, the source defining aperture is probably large enough that its diffraction effects on light reaching perimeter i can be neglected. One can approximately account for the geometric effects of the source defining aperture on the diffraction effects of nonlimiting aperture i by treating the recessed disk of radius R_{BB} as an extended-area source when computing $\alpha_{G,i}(\lambda)$. Note that this does not invalidate our previous use of the boundary diffraction wave formulation, even though there is not always a line of sight between $\mathbf{r}_s$ and $\mathbf{r}_d$. Validity of the boundary diffraction wave formulation depends on whether the proverbial point $\mathbf{l}$ on the perimeter of nonlimiting aperture i approaches the said line of sight, and it does not do so when the nonlimiting aperture is illuminated through the source defining aperture.

If the perimeter of nonlimiting aperture i is fully illuminated, a scaling factor w_i can approximately account for diffraction effects of the source defining aperture on the diffraction effects of nonlimiting aperture i: We would obtain $\alpha_{G,i}(\lambda) = w_i\epsilon_{G,i}(\lambda)$. We can let $G_i(r_i)$ denote the factor that multiplies the total power that passes through nonlimiting aperture i because of diffraction. If a nonlimiting aperture perimeter is fully illuminated, it is easy to estimate $G_i(r_i)$ by use of published formulas.[4] The scaling factor w_i is related to the irradiance at the aperture perimeter, and we obtain the following relation, from which we can deduce w_i:

$$\mathrm{d}[\pi r_i^2 G_i(r_i)]/\mathrm{d}r_i = 2\pi r_i w_i. \tag{27}$$

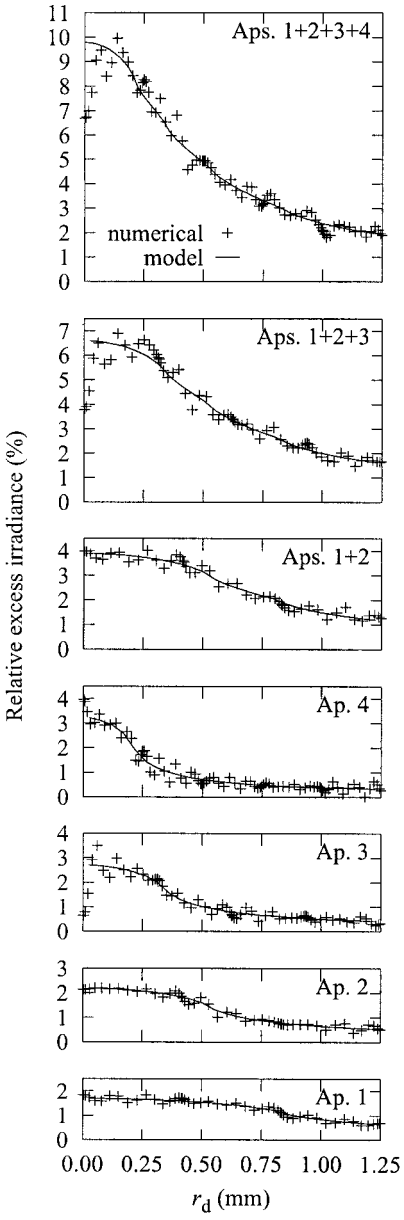

Fig. 2. Diffraction-induced relative excess irradiance for various combinations of apertures (Aps.) present in a setup studied by Boivin. Crosses indicate numerical results, and the solid curves indicate results given by the present model.

In two extreme cases with all N nonlimiting aperture perimeters fully illuminated, it can be especially easy to estimate the w_i scaling factors. In one limit, if the source defining aperture is large and its diffraction losses are small, we obtain $w_i \approx 1$ and $\epsilon(\lambda) \approx \epsilon_L(\lambda) + \epsilon_G(\lambda)$. In the opposite limit, if the source defining aperture is small and is illuminated from within a large solid angle, we obtain $w_i \approx F_L(\lambda)$ and $F(\lambda) \approx F_L(\lambda)F_G(\lambda)$.

4. Test of the Formula

As an example for testing the framework presented here, one can consider an optical setup used by Boivin to study diffraction effects.[11] The full setup is spec-

Table 2. Diffraction Effects for the Setup Used by Boivin

Aperture	Model $\epsilon_G(\lambda)$	Numerical $\epsilon_G(\lambda)$	Measured $\epsilon_G(\lambda)$
1	0.0110	0.0109	0.0105(5)
2	0.0095	0.0095	
3	0.0077	0.0076	
4	0.0057	0.0058	
1 + 2	0.0205	0.0206	0.0206(5)
1 + 2 + 3	0.0282	0.0283	0.0295(5)
1 + 2 + 3 + 4	0.0338	0.0340	0.0352(5)

ified in Table 1, although the four 7-mm-diameter nonlimiting apertures were not always present. In Ref. 6 and this paper the diffraction for $\lambda = 0.58$ μm is studied, which was the effective wavelength for the actual source–detector combination. In Ref. 6 theoretical diffraction losses due to the source defining aperture and gains due to the nonlimiting apertures were reported. In agreement with ideas raised here, it was found that diffraction effects of the nonlimiting apertures would arise from their being geometrically illuminated through the source defining aperture.

Carrying out the integration implied by Eq. (19), the diffraction effects on irradiance of the nonlimiting apertures can be modeled by approximation (18) on an aperture-by-aperture basis. Figure 2 shows the results found from Eq. (19), with the integrand computed according to approximation (18). In this setup, there are no geometric blocking effects, so that one always has nonzero $F_i(\mathbf{r}_s, \mathbf{r}_d)$ and $N_i(\mathbf{r}_s, \mathbf{r}_d)$. Figure 2 also shows the numerically calculated diffraction effects on irradiance based on a full numerical diffraction calculation in the spirit of approximation (2) by use of the Fresnel approximation and the methods of Ref. 6. It is gratifying that the results of approximation (18) and Eq. (19) agree so well with the full numerical results. Oscillations found in the present numerical results that are for monochromatic light vanish nearly altogether for a sufficiently broadband source, such as a Planck source. Numerical results for $F_G(\lambda)$ and the corresponding model results are shown in Table 2 along with available experimental results of Boivin (with $k = 1$ standard uncertainties being quoted). The agreement between the two types of theoretical result should be noted.

As a second example, Table 3 specifies a setup similar to one used to calibrate a blackbody source. Here we consider a fictitious blackbody with $R_s = 0.1$, 0.4, and 2.25 mm, and we consider effects of all three nonlimiting apertures, which are fully illuminated for the two smaller values of R_s but partially illuminated for the largest value.

Table 3. Specifications of a Plausible Blackbody (BB) Calibration Setup

Aperture	Radius (mm)	z Coordinate (mm)
BB cavity opening	$R_{BB} = 3$	$-d_{BB} = -15$
Source	$R_s = 0.1, 0.4, 2.25$	$(z_s = 0)$
1	$r_1 = 8.22$	$z_1 = 100.79$
2	$r_2 = 9.49$	$z_2 = 146.33$
3	$r_3 = 8.23$	$z_3 = 163.50$
Detector	$R_d = 10.33$	$z_d = 265.68$

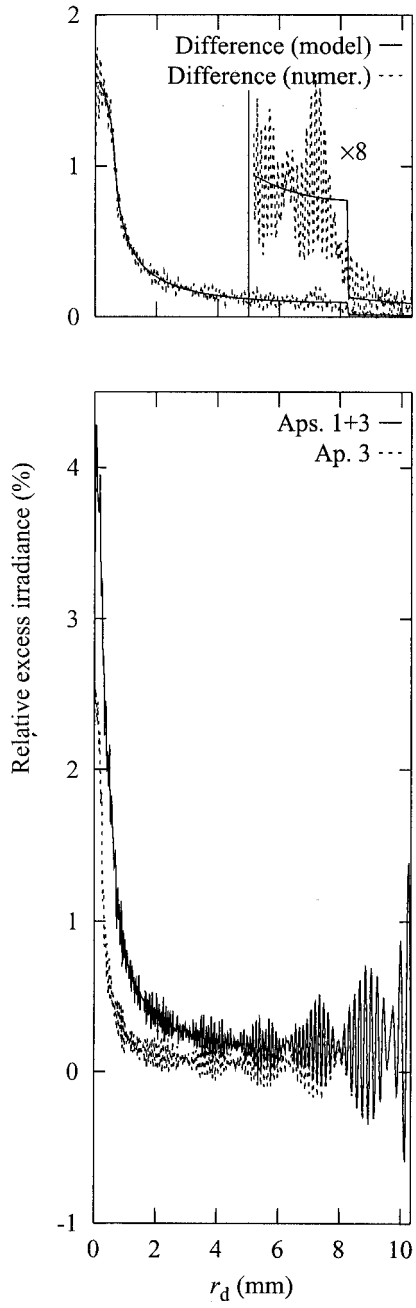

Fig. 3. Diffraction-induced relative excess irradiance for two combinations of apertures (Aps.) in a plausible setup for a blackbody calibration and the difference between the results (top panel). The solid curve indicates model results, and the dashed curve indicates numerical (numer.) results.

Figure 3 shows diffraction effects on spectral irradiance for 5-μm wavelength light and $R_s = 0.4$ mm for cases of only aperture 3 versus apertures 1 and 3. The difference is also shown and compared with model results expected from $w_1 F_1(\mathbf{r}_s, \mathbf{r}_d)$ and $w_1 N_1(\mathbf{r}_s, \mathbf{r}_d)$. The agreement appears reasonable on

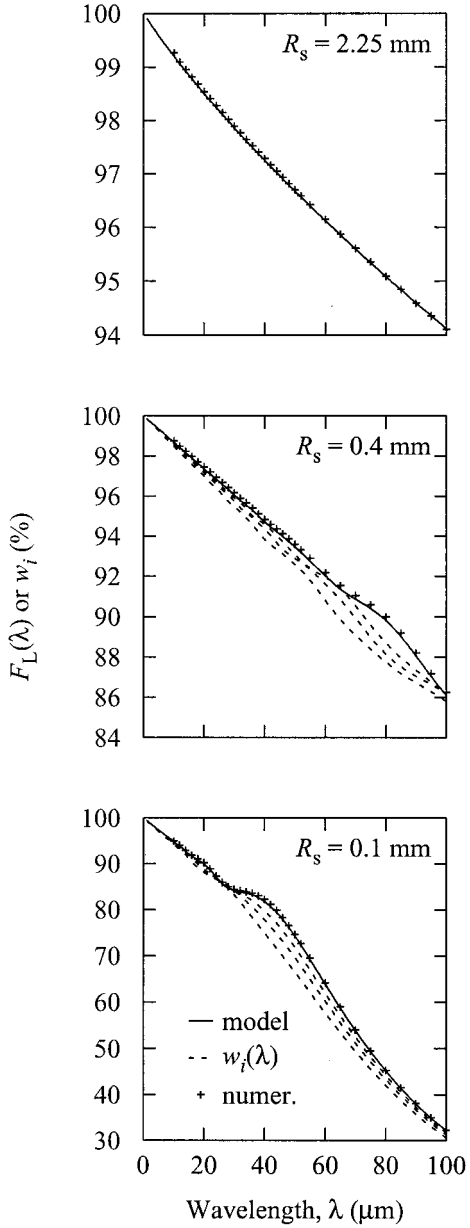

Fig. 4. Diffraction losses and (for two smaller R_s values) w_i parameters versus wavelength λ (dashed curves) for plausible setup for a blackbody calibration. The solid curves indicate model results and crosses indicate numerical (numer.) results.

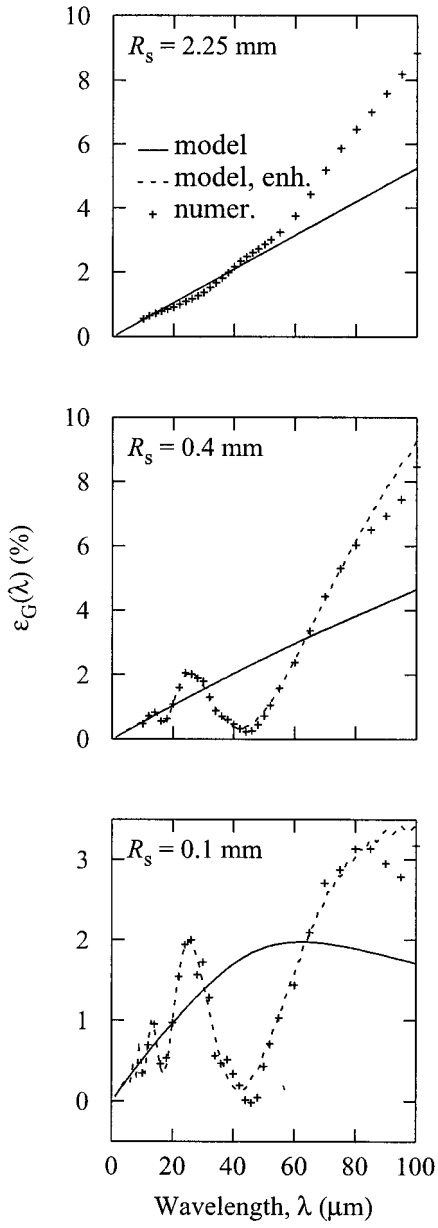

Fig. 5. Relative excess spectral power due to diffraction for a plausible setup for a blackbody calibration versus wavelength λ. The solid curves represent model results and crosses indicate numerical (numer.) results. Where appropriate, the model results were enhanced (enh.) as discussed in the text to give the dashed curves.

the basis of visual inspection, with oscillations in numerical results tending to be smaller for broadband light or shorter wavelengths. The oscillations also tend to be smaller for larger R_s and vice versa, which is a trend in many optical setups. This means that the model results presented in this paper tend to be better in cases in which numerical results are more demanding. Closer examination shows that the added irradiance due to aperture 1 bears the signature of Fresnel edge diffraction at the edge of aperture 3, which corroborates the notion of geometric blocking in the limit of short wavelengths.

For R_s = 0.1 mm and R_s = 0.4 mm, and all three nonlimiting apertures present, Eq. (27) and the results of Ref. 4 lead to values of $F_L(\lambda)$ and w_i that are plotted in Fig. 4. Likewise, Fig. 5 shows numerical results (crosses) and results of the present model (solid curves) for $\epsilon_G(\lambda)$. Clearly, there are oscillations that are missing in the model. When considering total power, the effects of these oscillations are largely self-canceling for a sufficiently broadband source, such as a thermal source. These oscillations arise, in particular, when an aperture is fully illuminated and, in an analogous fashion, fully viewed by the detector. In such cases, one can easily replace the model's $\epsilon_{G,i}(\lambda)$ with a more precise $\epsilon_{G,i}(\lambda)$ using

formulas such as those found in Ref. 4, which leads to the dashed curves shown in Fig. 5. Conversely, the present model can play the role of such formulas when an aperture is not fully illuminated and viewed. For $R_s = 2.25$ mm, the nonlimiting apertures are not all fully illuminated. Hence, we computed each $\alpha_{G,i}(\lambda)$ by treating the 3-mm-radius cavity opening as an extended-area source; values of the w_i parameters are not shown in the top panel of Fig. 4; and there is no dashed curve in the top panel of Fig. 5.

5. Closing Remarks

The potential value of this study is threefold. It can complement full numerical calculations, especially at small wavelengths. It can also help one check the validity of the numerical calculations. Finally, it may provide a good combination of computational accuracy and efficiency. Because this study is presented as a means of preventing larger, more accurate numerical calculations, the uncertainty of the results obtained can be characterized as follows. Differences between the results given by the formula that is developed and more complete numerical treatments provide an approximate example of the formula's accuracy. However, the overall uncertainty of the results obtained also has a component related to use of Kirchhoff diffraction theory, which is approximate, an issue beyond the scope of this paper that remains to be better studied.

References

1. W. R. Blevin, "Diffraction losses in radiometry and photometry," Metrologia **6,** 39–44 (1970).
2. W. H. Steel, M. De, and J. A. Bell, "Diffraction corrections in radiometry," J. Opt. Soc. Am. **62,** 1099–1103 (1972).
3. L. P. Boivin, "Diffraction corrections in radiometry: comparison of two different methods of calculation," Appl. Opt. **14,** 2002–2009 (1975).
4. E. L. Shirley, "Revised formulas for diffraction effects with point and extended sources," Appl. Opt. **37,** 6581–6590 (1998).
5. E. L. Shirley, "Fraunhofer diffraction effects on total power for a Planckian source," J. Res. Natl. Inst. Stand. Technol. **106,** 775–779 (2001).
6. E. L. Shirley and M. L. Terraciano, "Two innovations in diffraction calculations for cylindrically symmetrical systems," Appl. Opt. **40,** 4463–4472 (2001).
7. K. Miyamoto and E. Wolf, "Generalization of the Maggi-Rubinowicz theory of the boundary diffraction wave—Part I," J. Opt. Soc. Am. **52,** 615–625 (1962).
8. K. Miyamoto and E. Wolf, "Generalization of the Maggi-Rubinowicz theory of the boundary diffraction wave—Part II," J. Opt. Soc. Am. **52,** 626–637 (1962).
9. J. B. Keller, "Geometrical theory of diffraction," J. Opt. Soc. Am. **52,** 116–130 (1962).
10. M. Born and E. Wolf, *Principles of Optics*, 7th ed. (Cambridge U. Press, Cambridge, UK, 1999), pp. 484–499.
11. L. P. Boivin, "Reduction of diffraction errors in radiometry by means of toothed apertures," Appl. Opt. **17,** 3323–3328 (1978).